

Tartu Ülikool  
Loodus- ja täppisteaduste valdkond  
Matemaatika ja statistika instituut

Kaisa Komp

## **Pikkuse päritavuskoefitsiendi hindamine**

Matemaatilise statistika eriala  
Bakalaureusetöö (9 EAP)

Juhendajad Märt Möls,  
Reedik Mägi

Tartu 2017

## Pikkuse päritavuskoefitsiendi hindamine

Inimesed on üksteisest erinevad. Need erinevused on tingitud nii geneetilistest kui ka mitte geneetilistest ehk keskkonnateguritest. Päritavuskoefitsiendi hindamine aitab välja selgitada, kas inimestevahelised varieeruvad tunnused, näiteks pikkus, kaal või tõenäosus haigestuda erinevatesse haigustesse, on pigem seletatavad keskkonnast või geenidest tuleneva erinevusega. Käesoleva bakalaureusetöö eesmärk on leida Eestis elavate inimeste pikkuse päritavuskoefitsiendi hinnang erinevatel meetoditel. Kordajad on hinnatud valemiga, kus konkreetne sugulusside on teada, lineaarse mudeliga ja klassikalise kaksikute meetodiga. Tartu Ülikooli Geenivaramu andmete analüüsist selgub, et erinevate viiside rakendamine võib anda erisuguseid vastuseid. Lisaks saab kinnitust teadmine, et inimese pikkuse päritavuskoefitsient on kõrge.

**Märksõnad:** päritavus, inimese pikkus, lineaarne mudel, kaksikute meetod

CERCS teaduseriala: statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika (P160)

## Estimating heritability of height

People look different from each other. These dissimilarities are caused by genetic and nongenetic (environmental) factors. Heritability coefficient helps us find out if variable characteristics such as height, weight and probability of developing different diseases are possible to explain by environmental or genetic factors. The aim of this thesis is to estimate the heritability of height in Estonian residents by using different methods. The heritability is estimated by using a formula when specific relationship is known, and by using a linear model and twin method. It appears by the analysis of the data from the Estonian Genome Center, that using different methods can result in different answers. In addition, it will be confirmed that human height is mostly influenced by genetic factors.

**Keywords:** heritability, human height, linear model, twin method

CERCS research specialisation: statistics, operation research, programming, actuarial mathematics (P160)

# Sisukord

<b>Sissejuhatus</b>	<b>5</b>
<b>1 Töös kasutatud geneetika põhimõisted</b>	<b>7</b>
<b>2 Päritavuskoefitsiendi hindamine pidevate tunnuste korral</b>	<b>9</b>
2.1 Päritavuskoefitsiendi hindamine õvede või vanem-laps paaride korral .	15
2.2 Päritavuskoefitsiendi hindamine lineaarse mudeliga . . . . .	16
<b>3 Kaksikute meetod</b>	<b>17</b>
<b>4 Pikkuse päritavuskoefitsiendi hindamine</b>	<b>20</b>
4.1 Andmestike kirjeldus . . . . .	20
4.2 Pikkuse päritavuskoefitsiendi hinnang õvede või vanem-laps paaride korral . . . . .	22
4.3 Pikkuse päritavuskoefitsiendi hinnang lineaarse mudeliga . . . . .	23
4.4 Pikkuse päritavuskoefitsiendi hinnang kaksikute meetodil . . . . .	25
<b>5 Päritavuskoefitsiendi tõlgendamisest</b>	<b>26</b>
<b>Kokkuvõte</b>	<b>28</b>
<b>Kirjandus</b>	<b>29</b>
<b>LISAD</b>	<b>32</b>
Lisa 1. Meeste ja naiste pikkuse sõltuvust kirjeldavate mudelite võrdlus . .	32
Lisa 2. Meeste pikkuse sõltuvus sünniaastast . . . . .	33
Lisa 3. Naiste pikkuse sõltuvus sünniaastast . . . . .	34

## Sissejuhatus

Miks osal inimestel esineb mingi haigus ja teistel mitte? Kas see, et mõned inimesed on teistest pikemad, on peamiselt tingitud geenidest või keskkonnast? Sellistele küsimustele otsib vastuseid geneetiline epidemioloogia. Enamik fenotüübilisi tunnuseid ja haiguseid kujunevad nii erinevate geenide ja keskkonnamõjude kui ka nende omavaheliste interaktsioonide toimetel. Alles siis, kui on teada, et uuritaval tunnusel või haigusel on päritav komponent, on mõistlik hakata otsima sellega seotud gene.

Eelnevatest uuringutest on teada, et pikkuse päritavuskordaja on kõrge. Käesoleva bakalaureusetöö eesmärk on leida Eestis elavate inimeste pikkuse päritavuskoefitsient erinevatel meetoditel. Kordaja hindamiseks on kasutatud valemit, kus konkreetne sugulusside on teada, lineaarset mudelit ja kaksikute meetodit. Analüüsimisel on kasutatud Tartu Ülikooli Geenivaramu andmeid. Geenivaramuga liitunud inimesed on täisealised ning nõustunud oma andmete kasutamisega teadustööks. Antud töö on läbi viidud Tartu Ülikooli inimuuringute eetika komitee loa 234/T-12 raames.

Bakalaureusetöö koosneb järgmistest osadest. Esimeses peatükis defineeritakse töös kasutatud geneetika mõisteid. Teises osas on kirjeldatud, kuidas hinnatakse päritavuskoefitsient pidevate tunnuste korral. Lisaks on välja toodud valem, mille järgi hinnatakse päritavust övede või vanem-laps paaride korral. Samas on seletatud, kuidas saab ka lineaarset mudelit päritavuskordaja hindamisel kasutada. Kolmas peatükk annab ülevaate päritavuse hindamisest kaksikute meetodit kasutades. Neljandas peatükis on rakendatud eelnevalt mainitud viise pikkuse päritavuskordaja hindamiseks Tartu Ülikooli Geenivaramu andmete põhjal. Viimaks toob autor välja põhilised vead, mis esinevad koefitsiendi tõlgendamisel ning seletab, kuidas sellest paremini aru saada.

Töö on koostatud kasutades tekstitöötlusprogrammi Latex. Andmete analüüs on läbi viidud rakendustarkvara R-i abil.

Autor tänab töö juhendajaid Märt Mölsi ja Reedik Mägi huvitava teema, abistavate kommentaaride ja nõuannete eest. Samuti on tänuväärne nende kannatlikkus, ajaline panus ja abi andmete hankimisel.

# 1 Töös kasutatud geneetika põhimõisted

Järgnev lõik põhineb A. Heinaru (2012) õpikul *Geneetika*. Kõik organismid koosnevad rakkudest, mille tuumas asuvad kromosoomid. Inimese rakutuumas asub 23 paari kromosoomi. Iga paari korral on päritud üks kromosoom emalt ja teine isalt. Geenid ehk geneetilise informatsiooni kogumid asuvad kromosoomides ning nad määravad ühe või mitme tunnuse kujunemise. Igas kromosoomide paaris on järjestikku samad geenid. Alleelideks nimetatakse geeni erivorme. Geen sisaldab kahte alleeli, millest üks on päritud ühelt ja teine teiselt vanemalt. Lookus on koht kromosoomis, kus paikneb mingi geen. Fenotüübiks nimetatakse organismi tunnuseid (näiteks pikkus või vererõhk), mis on kujunenud genotüübi ja keskkonnategurite koosmõjul. Genotüüp on indiviidi geneetiline informatsioon või lookuste alleelne koosseis. Keskkonnamõjuks nimetatakse fenotüübi osa, mis ei ole organismile pärandunud tema vanematelt, vaid on tingitud välistest teguritest.

Päritavus on kvantitatiivse tunnuse muutlikuse osa, mis tuleneb indiviidide geneetilisest erinevusest mingis kindlas populatsioonis (Heinaru 2012: 893). Tunnuse päritavust iseloomustab päritavuskoeffitsient. Antud bakalaureusetöös on käsitletud kitsatähenduslikku päritavuskoeffitsienti, sest erinevate geenide vastastikune mõju ja alleelide domineerimisest tingitud mõju ei pärandu edasi, mistõttu oleme arvestanud neid keskkonnamõju hulka (Teinberg 1978: 271; Kaart 2001: 53). Kitsatähenduslik päritavuskoeffitsient, edaspidi päritavuskoeffitsient või päritavuskordaja, on defineeritud alleelide aditiivse toime dispersiooni ( $D[A_i]$ ) ja fenotüübi varieeruvuse ( $D[F_i]$ ) suhtena (Kaart 2012a: 3):

$$h^2 = \frac{D[A_i]}{D[F_i]}. \quad (1)$$

Alleelide aditiivne toime, edaspidi ka geneetiline mõju, on geenide summaarne mõju mingile uuritavale tunnusele (Heinaru 2012: 965). Päritavuskoeffitsiendi minimaalne väärtus on null, mis tähendab, et mingis kindlas populatsioonis on vaadeldava tunnuse erinevus seletatav indiviidide keskkonna erinevusega ning kordaja maksimaalne väärtus on üks, mis tähendab, et populatsioonis on uuritava tunnuse varieeruvus

põhjustatud isikute geneetilisest erinevusest (Heinaru 2012: 893).

Päritavuskordaja leidmiseks on vaja teada inimestevahelist sugulust, mida mõõdab suguluskoefitsient. Kaks inimest on omavahel suguluses, kui üks on teise järeltulija, või neil on vähemalt üks ühine esivanem (Nyquist 1990: 1). Defineerime  $i$ -nda ja  $j$ -nda indiviidide vahelise suguluskoefitsiendi  $r_{ij}$  alljärgnevalt:

$$r_{ij} = 0.5P(IBD = 1) + P(IBD = 2),$$

kus  $P(IBD = 1)$  on tõenäosus, et ühe inimese juhuslikult valitud lookuses leidub üks selline alleel, mis on päritolult identne teise inimese samas lookuses oleva ühe alleeliga, ja  $P(IBD = 2)$  on tõenäosus, et kahe isiku samas lookuses olevad mõlemad alleelid on omavahel päritolult identsed (Sham 1998: 208). Kahe inimese samas lookuses olevad alleelid on päritolult identsed (IBD), kui need on saadud ühiselt esivanemalt (Kaart 2012b: 1). Suguluskoefitsiendi  $r_{ij}$  hinnanguks on kahe indiviidi päritolult identsete alleelide osakaal (Kaart 2012b: 4). Antud suuruse minimaalne väärtus on 0, mis tähendab, et kaks isikut ei ole omavahel sugulased, ja maksimaalne väärtus on 1, mis tähendab, et kaks isikut on omavahel geneetiliselt identsed (Beebe, Rowe 2004: 103).



## 2 Päritavuskoefitsiendi hindamine pidevate tunnuste korral

Olgu  $i$ -nda indiviidi fenotüüp defineeritud järgnevalt:

$$F_i = \mu + A_i + \mathcal{E}_i, \quad (2)$$

kus  $\mu$  on uuritava tunnuse keskmine,  $A_i$  on  $i$ -nda isiku alleelide aditiivne mõju fenotüübi muutlikusele ja suurus  $\mathcal{E}_i$  tähistab keskkonnamõju  $i$ -nda indiviidi fenotüübile.

Eeldame, et geneetiline mõju ja keskkonnamõjude keskväärtused on võrdsed nulliga ehk

$$E[A_i] = 0 \text{ ja } E[\mathcal{E}_i] = 0 \quad \forall i \text{ korral,}$$

seega fenotüübi keskväärtus on võrdne:

$$\begin{aligned} E[F_i] &= E[\mu + A_i + \mathcal{E}_i] \\ &= E[\mu] + E[A_i] + E[\mathcal{E}_i] \\ &= \mu. \end{aligned}$$

Lisaks eeldame, et keskkonnamõju on sõltumatu geneetilisest mõjust ( $\mathcal{E}_i \perp A_i$ ), siis populatsiooni uuritava tunnuse dispersioon avaldub:

$$\begin{aligned} D[F_i] &= D[\mu + A_i + \mathcal{E}_i] \\ &= D[\mu] + D[A_i] + D[\mathcal{E}_i] \\ &= D[A_i] + D[\mathcal{E}_i]. \end{aligned}$$

Edasi leiame kahe isiku fenotüüpide vahelise kovariatsiooni. Kui  $i$ -nda ja  $j$ -nda indiviidi keskkonnamõjud on omavahel sõltumatud ning geneetiline mõju on sõltumatu

keskkonnamõjust, siis kovariatsioon kahe isiku fenotüübi vahel on:

$$\begin{aligned}\text{Cov}[F_i, F_j] &= \text{Cov}[\mu + A_i + \mathcal{E}_i, \mu + A_j + \mathcal{E}_j] \\ &= \text{Cov}[A_i, A_j].\end{aligned}$$

Vanematelt pärandub lapsele üks alleel emalt ja teine alleel isalt. Seega on võimalik aditiivne geneetiline mõju  $A_i$  kirja panna  $i$ -ndale inimesele pärandunud alleelide kaudu:

$$A_i = \sum_{K=1}^N (L_{iK}^E + L_{iK}^I),$$

kus suurused  $L_{iK}^E$  ja  $L_{iK}^I$  on vastavalt emalt ja isalt päritud  $K$ -nda lookuse mõju, mis on pärit jaotusest  $L_K$ . Suurus  $N$  tähistab geenide arvu, mis omavad mõju uuritavale fenotüübile.

Oletame, et ema ja isa ei ole omavahel sugulased. Seetõttu on emalt päritud  $K$ -nda lookuse mõju sõltumatu isalt päritud  $K$ -nda lookuse mõjust ( $L_{iK}^E \perp L_{iK}^I \ \forall K$ ). Aheldunud geenideks nimetatakse samas kromosoomis lähestikku paiknevaid geene, mis päranduvad koos (Heinaru 2012: 140). Eeldame edaspidi, et lookused asetsevad teineteisest piisavalt kaugel ja ei pärandu koos ehk kromosoomis ei ole aheldunud geene. Järelikult on samas kromosoomis paiknevad lookused omavahel sõltumatud ( $L_{iK}^E \perp L_{iL}^E, L_{iK}^I \perp L_{iL}^I \ \forall K, L = 1 \dots N, K \neq L$  korral). Lisaks eeldame, et lookuse mõju keskvääratus on võrdne nulliga ehk

$$E[L_{iK}^I] = 0 \text{ ja } E[L_{iK}^E] = 0 \ \forall i, K \text{ korral.} \quad (3)$$

Kui  $i$ -nda ja  $j$ -nda indiviidide keskkonnamõjud on omavahel sõltumatud ning aditiivne geneetiline mõju on sõltumatu keskkonnamõjust, siis kovariatsioon kahe isiku

fenotüübi vahel on:

$$\begin{aligned}
\text{Cov}[F_i, F_j] &= \text{Cov}[\mu + A_i + \mathcal{E}_i, \mu + A_j + \mathcal{E}_j] \\
&= \text{Cov} \left[ \mu + \sum_{K=1}^N (L_{iK}^E + L_{iK}^I) + \mathcal{E}_i, \mu + \sum_{L=1}^N (L_{jL}^E + L_{jL}^I) + \mathcal{E}_j \right] \\
&= \sum_{K=1}^N \sum_{L=1}^N \text{Cov} [L_{iK}^E + L_{iK}^I, L_{jL}^E + L_{jL}^I] \\
&= \sum_{K=L} \text{Cov} [L_{iK}^E + L_{iK}^I, L_{jL}^E + L_{jL}^I] + \sum_{K \neq L} \text{Cov} [L_{iK}^E + L_{iK}^I, L_{jL}^E + L_{jL}^I] \\
&= \sum_{K=1}^N \text{Cov} [L_{iK}^E + L_{iK}^I, L_{jK}^E + L_{jK}^I].
\end{aligned}$$

Liidetavad, kus  $K \neq L$ , on olukord, kus vaatleme mõlema isiku fenotüübi erinevaid lookuseid. Kuna eeldasime, et lookused pole aheldunud, siis on erinevad lookused omavahel sõltumatud. Järelikult on kovariatsioonid lookuste vahel võrdsed nulliga. Liidetavad, kus  $K = L$ , on olukord, kus vaatleme mõlema isiku fenotüübi samu lookuseid.

Olgu  $P(IBD = 0)$  tõenäosus, et kahe indiviidi mõlemad alleelid ei ole päritolult identsed,  $P(IBD = 1)$  tõenäosus, et ühe inimese juhuslikult valitud lookuses lei-  
dub üks selline alleel, mis on päritolult identne teise inimese samas lookuses oleva  
ühe alleeliga ja  $P(IBD = 2)$  tõenäosus, et kahe isiku mõlemad alleelid on pärito-  
lult identsed. Avaldades fenotüüpide kovariatsiooni tinglike kovariatsioonide kaudu  
(Rudary 2009: 126),

$$\text{Cov} [X, Y] = \text{E} [\text{Cov} [X, Y | Z]] + \text{Cov} [\text{E} [X | Z], \text{E} [Y | Z]],$$

saame tulemuseks:

$$\begin{aligned}
\text{Cov} [F_i, F_j] &= \sum_{Z=0}^2 \left( P(IBD = Z) \left( \sum_{K=1}^N \text{Cov} [L_{iK}^E + L_{iK}^I, L_{jK}^E + L_{jK}^I | IBD = Z] \right) \right) \\
&\quad + \text{Cov} \left[ \text{E} \left[ \sum_{K=1}^N (L_{iK}^E + L_{iK}^I) | IBD \right], \text{E} \left[ \sum_{K=1}^N (L_{jK}^E + L_{jK}^I) | IBD \right] \right].
\end{aligned}$$

Eeldame, et inimese alleelide mõju fenotüübi muutlikusele ei sõltu sellest, kas ja kui palju on sellel inimesel päritolult identseid allele teiste isikutega ehk

$$L_{iK}^E + L_{iK}^I \perp IBD.$$

Sel põhjusel on fenotüüpide kovariatsiooni leidmisel liidetav

$$\text{Cov} \left[ \mathbb{E} \left[ \sum_{K=1}^N (L_{iK}^E + L_{iK}^I) | IBD \right], \mathbb{E} \left[ \sum_{K=1}^N (L_{jK}^E + L_{jK}^I) | IBD \right] \right]$$

võrdne nulliga, sest tehtud sõltumatuse eelduse tõttu

$$\begin{aligned} \mathbb{E} [L_{iK}^E + L_{iK}^I | IBD] &= \mathbb{E} [L_{iK}^E + L_{iK}^I] \\ &= 0. \end{aligned}$$

Viimane võrdus kehtib eelduse (3) põhjal. Seega

$$\begin{aligned} \mathbb{E} \left[ \sum_{K=1}^N (L_{iK}^E + L_{iK}^I) | IBD \right] &= \sum_{K=1}^N \mathbb{E} [L_{iK}^E + L_{iK}^I | IBD] \\ &= 0 \end{aligned}$$

ning  $\text{Cov}[0,0] = 0$ .

Leiame edasi kahe inimese fenotüüpide vahelise kovariatsiooni, arvestades eelnevat tulemust:

$$\begin{aligned} \text{Cov} [F_i, F_j] &= P(IBD = 0) \left( \sum_{K=1}^N \text{Cov} [L_{iK}^E + L_{iK}^I, L_{jK}^E + L_{jK}^I | IBD = 0] \right) \\ &\quad + P(IBD = 1) \left( \sum_{K=1}^N \text{Cov} [L_{iK}^E + L_{iK}^I, L_{jK}^E + L_{jK}^I | IBD = 1] \right) \\ &\quad + P(IBD = 2) \left( \sum_{K=1}^N \text{Cov} [L_{iK}^E + L_{iK}^I, L_{jK}^E + L_{jK}^I | IBD = 2] \right). \end{aligned}$$

Kahe indiviidi fenotüübi alleelid on omavahel sõltuvad, siis kui nad on päritolult identsed, niisiis ei saa ka kovariatsioon lookuste vahel võrdsuda nulliga. Kahe indiviidi fenotüübi alleelid on omavahel sõltumatud, siis kui nad ei ole päritolult identsed,

järelikult on kovariatsioon lookuste vahel võrdne nulliga.

Kui  $IBD = 0$ , siis kahe indiviidi mõlemad alleelid ei ole päritolult identsed, seega on need alleelid sõltumatud ja kovariatsioonid lookuste vahel on võrdsed nulliga:

$$\sum_{K=1}^N \text{Cov} [L_{iK}^E + L_{iK}^I, L_{jK}^E + L_{jK}^I | IBD = 0] = 0.$$

Kui  $IBD = 1$ , siis oletame, et identsed on emalt päritud alleelid ehk  $L_{iK}^E = L_{jK}^E$ . Teised alleelid ei ole päritolult identsed ning seetõttu on nad sõltumatud ja vastavad kovariatsioonid on võrdsed nulliga. Seega, kui  $IBD = 1$ , saame kovariatsiooni tulemuseks:

$$\begin{aligned} \sum_{K=1}^N \text{Cov} [L_{iK}^E + L_{iK}^I, L_{jK}^E + L_{jK}^I | IBD = 1] &= \sum_{K=1}^N \text{Cov} [L_{iK}^E + L_{iK}^I, L_{iK}^E + L_{jK}^I | IBD = 1] \\ &= \sum_{K=1}^N \left( \text{Cov} [L_{iK}^E, L_{iK}^E | IBD = 1] + \text{Cov} [L_{iK}^E, L_{jK}^I | IBD = 1] \right. \\ &\quad \left. + \text{Cov} [L_{iK}^I, L_{iK}^E | IBD = 1] + \text{Cov} [L_{iK}^I, L_{jK}^I | IBD = 1] \right) \\ &= \sum_{K=1}^N \text{Cov} [L_{iK}^E, L_{iK}^E | IBD = 1] = \sum_{K=1}^N DL_K. \end{aligned}$$

Viimane võrdus tuleneb kovariatsiooni omadusest. Eelnev arutelu ja tulemus ei muutu, kui päritolult identsed on teised alleelid ( $L_{iK}^E = L_{jK}^I$ ,  $L_{iK}^I = L_{jK}^E$  või  $L_{iK}^I = L_{jK}^I$  korral), sest  $K$ -nda lookuse alleelide mõjud on pärit samast jaotusest  $L_K$ , mille dispersioon on  $DL_K$ .

Kui  $IBD = 2$ , siis oletame, et identsed on nii emalt päritud alleelid kui ka isalt päritud alleelid ehk  $L_{iK}^E = L_{jK}^E$  ja  $L_{iK}^I = L_{jK}^I$ . Vastava kovariatsiooni tulemuseks

saame:

$$\begin{aligned}
\sum_{K=1}^N \text{Cov}[L_{iK}^E + L_{iK}^I, L_{jK}^E + L_{jK}^I | IBD = 2] &= \sum_{K=1}^N \text{Cov}[L_{iK}^E + L_{iK}^I, L_{iK}^E + L_{iK}^I | IBD = 2] \\
&= \sum_{K=1}^N \left( \text{Cov}[L_{iK}^E, L_{iK}^E | IBD = 2] + \text{Cov}[L_{iK}^E, L_{iK}^I | IBD = 2] \right. \\
&\quad \left. + \text{Cov}[L_{iK}^I, L_{iK}^E | IBD = 2] + \text{Cov}[L_{iK}^I, L_{iK}^I | IBD = 2] \right) \\
&= \sum_{K=1}^N \left( \text{Cov}[L_{iK}^E, L_{iK}^E | IBD = 2] + \text{Cov}[L_{iK}^I, L_{iK}^I | IBD = 2] \right) \\
&= \sum_{K=1}^N (D L_K + D L_K) = 2 \sum_{K=1}^N D L_K.
\end{aligned}$$

Vastus ei muutu, kui päritolult identsed on teised alleelid, sest K-nda lookuse alleelide mõjud on samast jaotusest  $L_K$ .

Arvestades eelnevaid kovariatsioone,  $IBD = 0$ ,  $IBD = 1$  ja  $IBD = 2$  korral, siis saame  $\text{Cov}[F_i, F_j]$  tulemuseks:

$$\begin{aligned}
\text{Cov}[F_i, F_j] &= P(IBD = 1) \sum_{K=1}^N D[L_K] + P(IBD = 2) 2 \sum_{K=1}^N D[L_K] \\
&= 2 \left( \frac{1}{2} P(IBD = 1) + P(IBD = 2) \right) \sum_{K=1}^N D[L_K] \\
&= 2 r_{ij} \sum_{K=1}^N D[L_K],
\end{aligned}$$

kus  $r_{ij}$  on  $i$ -nda ja  $j$ -nda indiviidide vaheline suguluskoefitsient ja  $2 \sum_{K=1}^N D[L_K] = D[A_i]$  on geenide aditiivsusest tingitud dispersioon. Kuna  $\text{Cov}[F_i, F_j] = r_{ij} D[A_i]$ , siis saame aditiivsusest tingitud dispersiooni avaldada:

$$D[A_i] = \frac{\text{Cov}[F_i, F_j]}{r_{ij}}. \quad (4)$$

Arvutades sel viisil aditiivsusest tingitud dispersiooni, peame jälgima, et suguluskordaja  $r_{ij}$  ei võrduks nulliga. Tulemust (4) kasutades saame päritavuskordaja arvutamiseks valemi:

$$h^2 = \frac{D[A_i]}{D[F_i]} = \frac{\text{Cov}[F_i, F_j]}{r_{ij} D[F_i]}. \quad (5)$$

Pisut teistsuguse aruteluga on jõutud samasuguse tulemuseni ka raamatus *Population Genetics* (Gillespie 2004: 144- 146).

## 2.1 Päritavuskoefitsiendi hindamine õvede või vanem-laps paaride korral

Olgu  $i$ -nda indiviidi fenotüüp defineeritud nagu valemis (2). Oletame, et  $i$ -ndas ja  $j$ -ndas isik on omavahel vanem ja järglane või samade vanemate lapsed. Teame, et suguluskoefitsient on õvede korral keskmiselt ning lapse ja tema vanema korral täpselt 0,5 (Hedrick 2011: 473). Kasutades tulemust (5) ja teadmist, et  $r_{ij} = 0,5$ , võib päritavuskordaja leida valemiga:

$$\begin{aligned} h^2 &= \frac{D[A_i]}{D[F_i]} \\ &= \frac{\text{Cov}[F_i, F_j]}{0,5D[F_i]}. \end{aligned}$$

Valimi põhjal saab leida  $\frac{\text{Cov}[F_i, F_j]}{D[F_i]}$  hinnangu  $\frac{\widehat{\text{Cov}[F_i, F_j]}}{\widehat{D[F_i]}}$ , sest fenotüübid on teada. Niisiis arvutatakse päritavuskoefitsient õvede või lapse ja tema vanema korral alljärgnevalt:

$$\begin{aligned} h^2 &= \frac{\widehat{\text{Cov}[F_i, F_j]}}{0,5\widehat{D[F_i]}} \\ &= \frac{\frac{1}{n-1} \sum_{k=1}^n (F_i - \overline{F_i})(F_j - \overline{F_j})}{0,5s^2}, \end{aligned} \tag{6}$$

kus  $n$  on paaride arv valimis. Suurus  $F_i$  on  $i$ -nda ema/isa või venna/õe fenotüüp ja  $F_j$  tähistab  $i$ -nda isiku lapse või venna/õe fenotüüpi. Vanemate või õvede keskmine on  $\overline{F_i}$  ja nendele vastavate laste või õvede keskmine on  $\overline{F_j}$ . Suurus  $s^2$  on valimi põhjal arvutatud fenotüübi dispersioon.

## 2.2 Pärítavuskoefitsiendi hindamine lineaarse mudeliga

Pärítavuskoefitsiendi leidmiseks üldjuhu korral hinnatakse lineaarse mudeli regressioonikordajat. Selle paremaks mõistmiseks korrutame pärítavuskordaja valemi (1) võrduse mõlemaid pooli  $r_{ij}$ -ga ja saame tulemuseks:

$$\frac{D[A_i]r_{ij}}{D[F_i]} = h^2 r_{ij}.$$

Varem selgus, et  $r_{ij}D[A_i] = \text{Cov}[F_i, F_j]$ . Kui Asendada  $r_{ij}D[A_i]$  ülaltoodud võrduses kahe isiku fenotüüpide vahelise kovariatsiooniga, siis

$$\frac{\text{Cov}[F_i, F_j]}{D[F_i]} = h^2 r_{ij}.$$

Saadud võrdust võib vaadelda kui ühte tundmatut parameetrit ( $h^2$ ) sisaldavat regressioonanalüüsi mudelit. Funktsioontunnuseks  $Y$  on  $\frac{\text{Cov}[F_i, F_j]}{D[F_i]}$  hinnang  $\frac{\widehat{\text{Cov}[F_i, F_j]}}{\widehat{D[F_i]}}$ , mis arvutatakse:

$$\frac{\widehat{\text{Cov}[F_i, F_j]}}{\widehat{D[F_i]}} = \frac{(F_i - \bar{F})(F_j - \bar{F})}{s^2},$$

kus  $s^2$  on fenotüübi dispersioon valimis,  $F_i$  ja  $F_j$  on vastavalt  $i$ -nda ja  $j$ -nda indiviidi fenotüüp ja  $\bar{F}$  on uuritava tunnuse keskmine. Argumenttunnuseks  $X$  on  $i$ -nda ja  $j$ -nda indiviidi suguluskoefitsient  $r_{ij}$ , mis on arvutatud indiviidide alleelide põhjal. Hinnatavaks parameetriks  $\beta$  on meie otsitav pärítavuskordaja ning saame lineaarse regressioonimudeli kujul

$$Y = \beta X + \varepsilon,$$

kus  $\varepsilon$  mudeli juhuslik viga.



### 3 Kaksikute meetod

Kõige populaarsem viis päritavuse hindamiseks on kaksikute meetod. Ühemunakaksikud ehk monosügootsed kaksikud on geneetiliselt identsed ehk suguluskoefitsient on võrdne ühega ning kahemunakaksikute ehk disügootsete kaksikute genotüübid on keskmiselt 50% ulatuses ühesugused, mis on sama nagu teistel samade vanemate lastel (Heinaru 2012: 894). Päritavuskordaja hindamine kaksikute meetodil põhineb ühe- ja kahemunaraku kaksikute võrdlemisel. Uuringutes eeldatakse, et monosügootseid kaksikuid ei kohelda sarnasemalt kui disügootseid kaksikuid (Sham 1998: 189). Kui ühemunakaksikud siiski on sarnasemad võrreldes kahemunakaksikutega, siis on see tingitud sellest, et ühemunakaksikutele on pärandunud rohkem sarnaseid genee kui kahemunakaksikutele (Sham 1998: 189).

Olgu  $i$ -nda kaksiku uuritav tunnus defineeritud järgnevalt:

$$F_i = \mu + A_i + \mathcal{E}_i + \mathcal{E},$$

kus  $\mu$  on uuritava tunnuse keskmine,  $A_i$  on  $i$ -nda kaksiku geneetiline mõju fenotüübi muutlikusele,  $\mathcal{E}_i$  on  $i$ -nda kaksiku keskkonnamõju fenotüübile ja  $\mathcal{E}$  on  $i$ -nda kaksiku ja tema venna või õe ühine keskkonnaefekt (näiteks samad vanemad ja toit) fenotüübile.

Järgnevalt leiame kaksikute fenotüüpide vahelise kovariatsiooni. Selleks eeldame, et geneetiline mõju on sõltumatu keskkonnamõjudest ( $A_i \perp \mathcal{E}_i$  ja  $A_i \perp \mathcal{E}$ ) ja kaksikute ühesugune elukeskkond on sõltumatu keskkonnatingimustest, mis ei ole ühine ( $\mathcal{E} \perp \mathcal{E}_i$ ).

Varem saime, et  $\text{Cov}[F_i, F_j] = \text{Cov}[A_i, A_j] = r_{ij}D[A_i]$ , kus  $r_{ij}$  on  $i$ -nda ja  $j$ -nda indiviide vaheline suguluskoefitsient ja  $D[A_i]$  on aditiivsusest tingitud dispersioon. Antud võrdus kehtib siis, kui kahe isiku keskkonnad on omavahel sõltumatud. Tulemust saab rakendada ka kaksikute fenotüüpide kovariatsioonide leidmisel, lisaks on vaja arvestada kaksikute ühesuguse keskkonnamõjuga.

Monosügootsete kaksikute vaheline suguluskoefitsient on võrdne ühega ja disügootsete kaksikute vaheline sugulus on võrdne 0,5-ga. Niisiis arvestades eeldusi ( $A_i \perp \mathcal{E}_i$ ,  $A_i \perp \mathcal{E}$  ja  $\mathcal{E} \perp \mathcal{E}_i$ ), saame ühe- ja kahemunaraku kaksikute fenotüüpide vaheliseks kovariatsiooniks vastavalt:

$$\begin{aligned}\text{Cov}[F_i, F_j]_{MS} &= \text{Cov}[\mu + A_i + \mathcal{E}_i + \mathcal{E}, \mu + A_j + \mathcal{E}_j + \mathcal{E}] \\ &= \text{Cov}[A_i, A_j] + \text{Cov}[\mathcal{E}, \mathcal{E}] \\ &= D[A_i] + D[\mathcal{E}]\end{aligned}$$

ja

$$\begin{aligned}\text{Cov}[F_i, F_j]_{DS} &= \text{Cov}[\mu + A_i + \mathcal{E}_i + \mathcal{E}, \mu + A_j + \mathcal{E}_j + \mathcal{E}] \\ &= \text{Cov}[A_i, A_j] + \text{Cov}[\mathcal{E}, \mathcal{E}] \\ &= \frac{1}{2}D[A_i] + D[\mathcal{E}].\end{aligned}$$

Kuna monosügootseid kaksikuid koheldakse samamoodi nagu disügootseid kaksikuid, siis ühine keskkonna varieeruvus on neil võrdne. Edasi võrdsustamegi ühe- ja kahemuna kaksikute keskkonnamõju varieeruvuse:

$$\text{Cov}[F_i, F_j]_{MS} - D[A_i] = \text{Cov}[F_i, F_j]_{DS} - \frac{1}{2}D[A_i],$$

millest järeldub, et

$$\frac{1}{2}D[A_i] = \text{Cov}[F_i, F_j]_{MS} - \text{Cov}[F_i, F_j]_{DS}.$$

Korrutades võrduse mõlemaid pooli kahega on tulemuseks geneetilise mõju dispersioon:

$$D[A_i] = 2\text{Cov}[F_i, F_j]_{MS} - 2\text{Cov}[F_i, F_j]_{DS}.$$

Asendame leitud aditiivsusest tingitud dispersiooni päritavuskoefitsiendi definitsioonis ja saame arvutusvalemi kordaja hindamiseks, kui valim koosneb ühe- ja kahe-

munaraku kaksikutest:

$$\begin{aligned}
h^2 &= \frac{D[A_i]}{D[F_i]} \\
&= \frac{2\text{Cov}[F_i, F_j]_{MS} - 2\text{Cov}[F_i, F_j]_{DS}}{D[F_i]} \\
&= 2 \left( \frac{\text{Cov}[F_i, F_j]_{MS}}{D[F_i]} - \frac{\text{Cov}[F_i, F_j]_{DS}}{D[F_i]} \right).
\end{aligned}$$

Uuritav tunnus kaksikute puhul ei erine kuidagimoodi teistest sama populatsiooni inimestest. Seetõttu on populatsiooni uuritava tunnuse varieeruvus võrdne kaksikute fenotüüpide varieeruvusega ehk  $D[F_i] = D[F_i]_{MS} = D[F_i]_{DS}$ . Niisiis saab päritavuskordajat arvutada ka korrelatsioonikordajate abil:

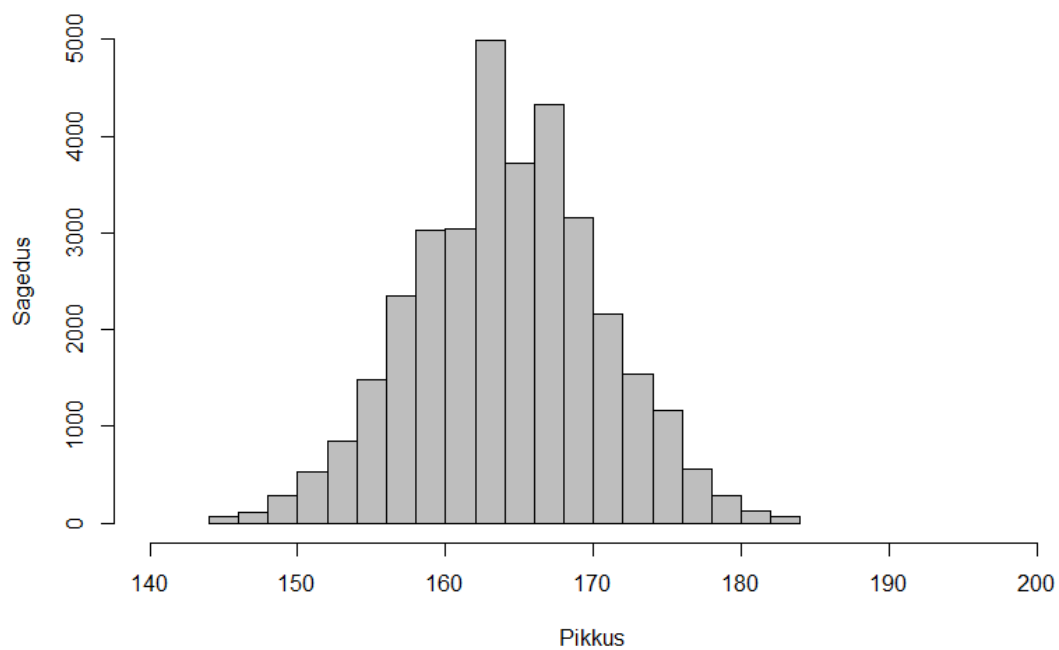
$$\begin{aligned}
h^2 &= 2 \left( \frac{\text{Cov}[F_i, F_j]_{MS}}{D[F_i]} - \frac{\text{Cov}[F_i, F_j]_{DS}}{D[F_i]} \right) \\
&= 2 \left( \frac{\text{Cov}[F_i, F_j]_{MS}}{D[F_i]_{MS}} - \frac{\text{Cov}[F_i, F_j]_{DS}}{D[F_i]_{DS}} \right) \\
&= 2(r_{MZ} - r_{DZ}).
\end{aligned}$$

Viimast võrdust nimetatakse ka Falconeri valemiks (Rijsdijk, Sham 2002: 120).

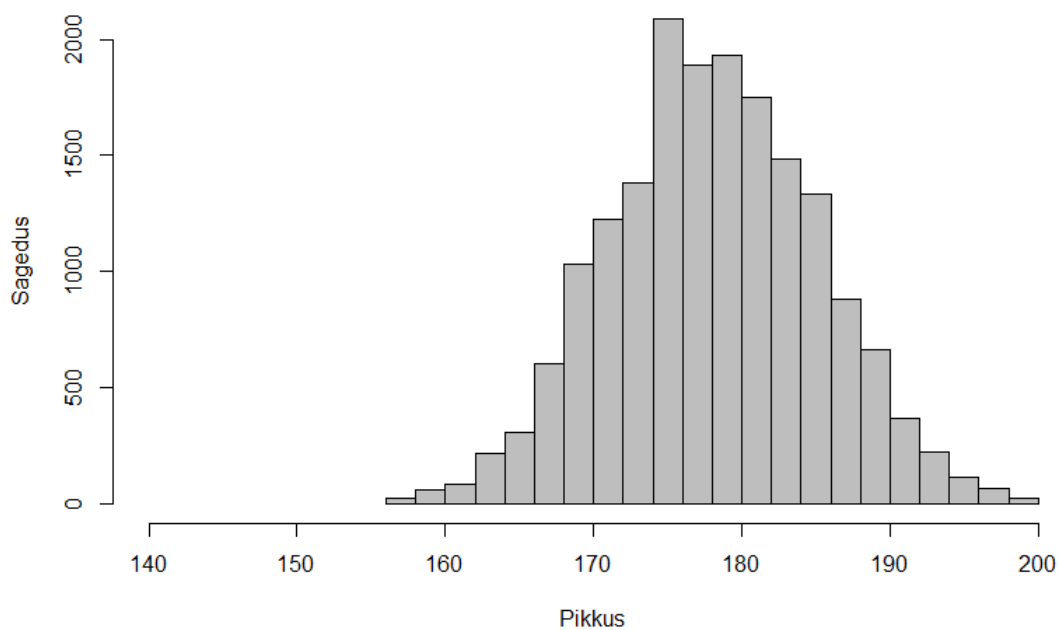
## 4 Pikkuse päritavuskoefitsiendi hindamine

### 4.1 Andmestike kirjeldus

Pikkuse päritavuskoefitsiendi hindamiseks on kasutatud Tartu Ülikooli Geenivaramu andmeid. Kasutatavaid andmestikke on kaks. Ühes andmestikus on 51 832 inimese pikkus, sugu ja sünniaasta ning teises inimestevahelised suguluskoefitsiendid. Analüüsimiseks on kasutatud 17 732 mehe ja 33 867 naise andmed. Eemaldatud on 69 inimese andmed, sest nende kohta ei ole pikkus teada. Erindite leidmisel on arvestatud kolme sigma reeglga- väärtused, mis jäävad vahemikust [keskmine-3-standardhälve; keskmine+3-standardhälve] välja, loetakse ebatavalisteks. Sel põhjusel ei võeta analüüsimisel arvesse veel 58 mehe ja 106 naise andmeid. Naiste ja meeste pikkuse jaotused on toodud vastavalt joonisel 1 ja 2. Joonistelt on näha, et mehed on keskmiselt pikemad kui naised, mistõttu ei ole need üheselt võrreldavad ning analüüsimisel tuleks seda arvestada. Lisaks on selgunud, et inimeste keskmine pikkus on aastatega kasvanud (NCD-RisC 2016). Seetõttu on päritavuskoefitsiendi hindamisel kasutatud standardiseeritud jääke, mis on saadud pikkuse ja sünniaasta sõltuvust kirjeldavast mudelist, mis on leitud meestel ja naistel eraldi. Sobivaima mudeli valimisest on lähemalt kirjutatud lisas 1. Meeste ja naiste pikkuse sõltuvust sünniaastast on kujutatud vastavalt lisas 2 ja 3.



Joonis 1. Naiste pikkuse jaotus



Joonis 2. Meeste pikkuse jaotus

Teises andmestikus on 24 098 547 inimestevahelist suguluskordajat. Kuna andmestik on suur, siis päritavuskordaja hindamisel ei pidanud autor vajalikuks niipaljude

mitte sugulaste lisamist valimisse. Seetõttu on mitte sugulased valitud valimisse juhuslikult. Pärast teadmata pikkuste ja erindite eemaldamist jääb järgi 204 496 suguluskordajat. Suguluskoefitsiendi võimalikeks väärtusteks on näiteks  $1, \frac{1}{2}, \frac{1}{4}$  jne (Hedrick 2011: 473). Kuna valimis olevad inimestevahelised sugulused on hinnatud ülegenoomselt alleelide põhjal, siis ei pruugi nende väärtused olla täpselt  $1, \frac{1}{2}, \frac{1}{4}$  jne. Seetõttu on loetud suguluskoefitsiendiks neid hinnatud suguluskordajaid, mis asuvad suguluskoefitsiendile lähedal. Näiteks suguluskoefitsiendi 0,25 hulka kuuluvad need hinnatud sugulused, mis asuvad poollõigus  $(0,188; 0,375]$ . Valim koosneb peamiselt inimeste paaridest, kes ei ole omavahel sugulased, mida on näha ka tabelist 1.

Tabel 1. Suguluskoefitsientide jaotus- ja sagedustabel

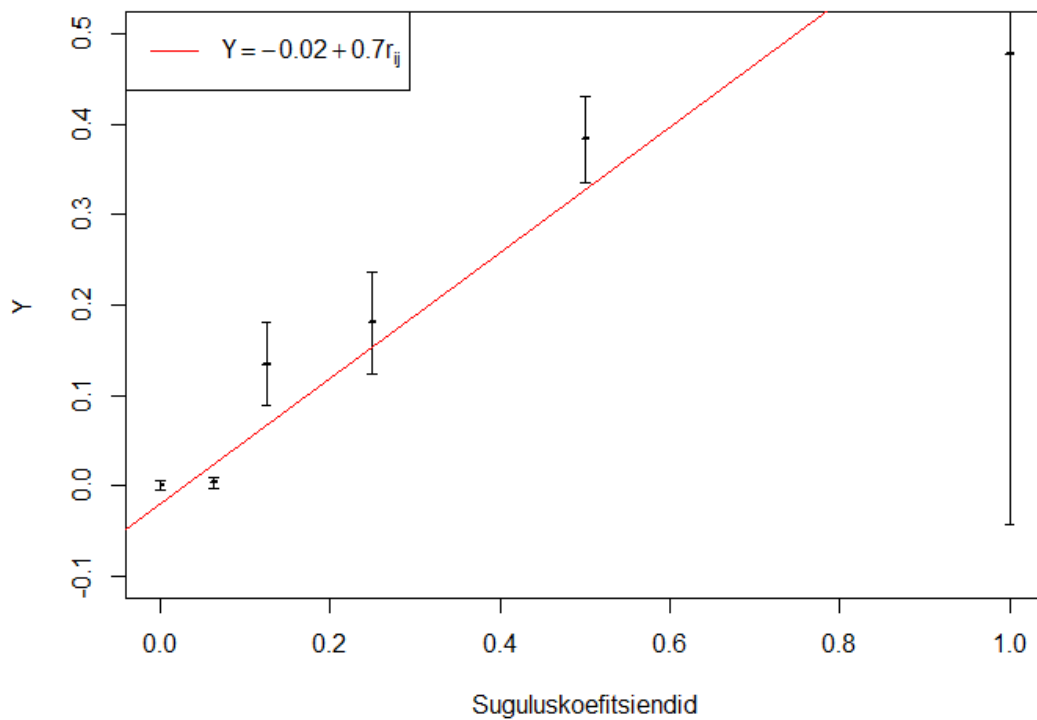
Suguluskoefitsient	0	0,0625	0,125	0,25	0,5	1
Sagedus	115 520	84 171	1 878	1 206	1 713	8
%	56,490	41,160	0,918	0,590	0,838	0,004

## 4.2 Pikkuse päritavuskoefitsiendi hinnang õvede või vanem-laps paaride korral

Andmestikus on 1713 lapse ja tema vanema või õvede paari (vt. tabel 1). Õvedeks või vanemaks ja lapseks loetakse need isikud, kellevahelised suguluskoefitsiendid jäävad poollõiku  $(0,375; 0,875]$ . Pikkuse päritavuskoefitsiendi hindamiseks saab kasutada valemit (6), kui valim koosneb õvedest või vanem-laps paaridest. Varasemates uuringutes on selgunud, et pikkuse päritavuskordaja on ligikaudu 0,80 (Silventoinen jt 2003; Visscher 2008: 489), mis on lähedane Geenivaramu andmete põhjal leitud tulemusele 0,81. Usaldusvahemikuks usaldusnivool 0,95 on saadud ligikaudu  $[0,73; 0,88]$ . Hinnatud on ka lineaarne regressioonimudel (vt. alampeatükk 2.2) ja leitud õvede või laps-vanema paaride korral pikkuse päritavuskoefitsient. Tulemuseks on saadud 0,80 ja 95%-line usaldusintervall on umbes  $[0,71; 0,90]$ . Kahel erineval viisil leitud hinnangud on peaaegu võrdsed.

### 4.3 Pikkuse päritavuskoefitsiendi hinnang lineaarse mudeliga

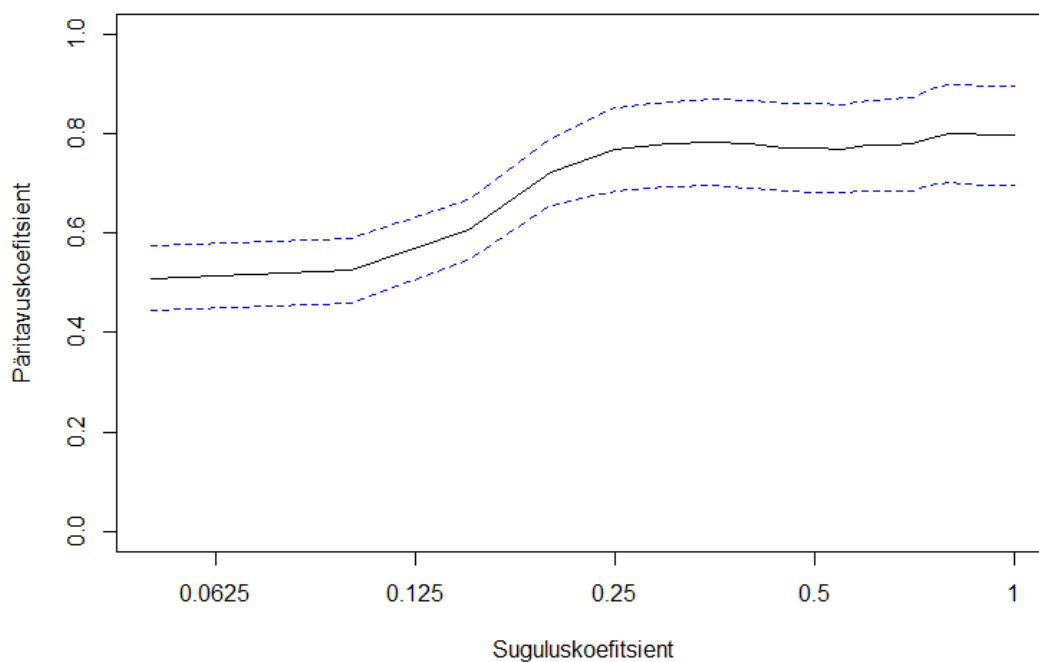
Järgnevalt hinnatakse pikkuse päritavuskoefitsient lineaarse mudeliga. Nüüd ei ole oluline konkreetne sugulusside ja arvesse võetakse kõik 204 496 suguluskordajat. Analüüsi tulemust on kujutatud joonisel 3. Lineaarse mudeli sirge tõus on 0,7, mis ongi hinnatav päritavuskordaja. Usaldusvahemikuks usaldusnivool 0,95 on saadud ligikaudu  $[0,61; 0,77]$ . Sel viisil leitud päritavuskoefitsiendi hinnang on 0,1 võrra väiksem võrreldes alampeatükis 4.2 saadud hinnanguga. Sellegipoolest saab mõlema tulemuse põhjal väita, et pikkuse varieeruvus on peamiselt tingitud geenide erinevusest.



Joonis 3. Tunnuse  $Y$  sõltuvus suguluskoefitsiendist. Sirge tõus tähistab päritavuskoefitsienti.

Päritavuskoefitsiendi sõltuvust suguluskoefitsiendist on kujutatud joonisel 4, millelt on näha, et mida suurem on inimestevaheline sugulus, seda suurem on päritavuskordaja hinnang. See võib viidata sellele, et omavahel lähedalt sugulaste elukeskkond ei ole nii erinev. Näiteks vanemad ja lapsed elavad koos ning nende toitumisharju-

mused on sarnased. Seetõttu on ka keskkonna erinevus väike ja proportsionaalselt päritavuse hinnang kõrge. Samas kaugelt sugulased ei ole samas leibkonnas, nende keskkonnaolud on erinevamad ning sel põhjusel on ka päritavuskordaja hinnang madalam. Lisaks tasub tähele panna eeldust, mis sai tehtud inimeste fenotüüpide vahelise kovariatsiooni leidmisel, et indiviidide keskkonnamõjud on omavahel sõltumatud. Ilmselge on, et see eeldus on rikutud, kui tegemist on näiteks õvedega või vanem-laps paaridega, sest neil on osa keskkonnast sarnasem kui juhuslikel inimestel ja seega keskkonnamõjud on sõltuvad. Selle tõttu võib ka alampeatükis 4.2 saadud hinnang olla veidi ülehinnatud.



Joonis 4. Päritavuse sõltuvus suguluskoefitsiendist. Sinine katkendlik joon tähistab 95%-list usaldusintervalli.



#### 4.4 Pikkuse päritavuskoeffitsiendi hinnang kaksikute meetodil

Andmestikus on 17 kaksikute paari, kellest kaheksa on ühemunaraku ja üheksa on kahemunaraku kaksikud. Ühemunakaksikuteks loetakse need inimesed, kellevahelised sugulused jäävad poollõiku  $(0,875; 1]$ . Kahemunakaksikute hulka arvestatakse need paarid, kellevaheline suguluskoeffitsient on poollõigus  $(0,375; 0,875]$  ja, kes on samal aastal sündinud. Autor on otsustanud arvutused läbi teha ka kaksikute meetodil, kuigi andmestik on üsna väike. Seetõttu sel viisil saadud tulemuste põhjal ei ole mõistlik mingeid järeldusi teha. Kuna päritavuse hindamine kaksikute meetodil eeldab, et monosügootseid kaksikuid koheldakse samamoodi nagu disügootseid kaksikuid, siis ei ole analüüsimisel arvesse võetud eri soost disügootseid kaksikuid. Sel põhjusel on eemaldatud kolm kahemunakaksikute paari. Pikkuse päritavuskoeffitsiendi hinnanguks kaksikute meetodil on saadud 0,95.

## 5 Päritavuskoefitsiendi tõlgendamisest

Esmapilgul võib tunduda päritavuse tähendusest aru saamine keeruline ning sageli interpreteeritakse seda valesti. Päritavuse paremaks mõistmiseks on autor esile toonud põhilised vead, mida tehakse päritavuskordaja tõlgendamisel.

Tihti arvatakse, et järeldusi saab teha indiviidi kohta (Haverkamp 2015:11). Näiteks päritavuskoefitsiendi väärtust 0,8 seletatakse nii, et 80% uuritava tunnuse väärtusest on tingitud geenidest. Seega on vale väita inimese pikkuse kohta, et 80% pikkusest on tingitud geenidest ja 20% on mõjutanud elukeskkond. Isegi päritavuskoefitsiendi definitsioonist (1) tuleb välja, et kordajat ei saa rakendada indiviidile, sest valemis on dispersioonid ning üksikisiku pikkus ei varieeru. Tegelikult tähendab päritavuskordaja suurus 0,8, et 80% fenotüübi varieeruvusest on seletatav geneetilise erinevusega. Niisiis on õige öelda, et inimeste pikkuste erinevust on põhjustanud 80% geenid ja 20% keskkond.

Teiseks väidetakse, et madal päritavus tähendab seda, et geenid ei mõjuta uuritava tunnuse kujunemist (Oldenbroek, Waaij 2014: 108). Madal päritavuskoefitsiendi suurus näitab seda, et uuritava tunnuse varieeruvus on peamiselt tingitud keskkonna erinevusest. Lihtsamalt öeldes tähendab see seda, et kui populatsioonis olevad inimesed üksteisest erinevad, siis on selle peamine põhjus inimeste erinevad kasvamise tingimused. Niisiis võib geenidel olla väga suur efekt mingi tunnuse kujunemisel, kuid kuna nende mõju on inimeste vahel samasugune, siis ei esine geneetilist varieeruvust, mistõttu on päritavuskordaja suurus väike (Haverkamp 2015:10).

Lisaks tasub meele pidada, et päritavuskoefitsiendi arvutatakse mingil ajahetkel ühes kindlas populatsioonis (Oldenbroek, Waaij 2014: 108). Seega vaadeldes ühe tunnuse päritavuskordajat, võivad koefitsientide väärtused olla erinevates populatsioonides erinevad. Näiteks on Miao- Xin Li (2004: 148) oma uurimisrühmaga saanud hiinlaste pikkuse päritavuskoefitsiendi hinnanguks ligikaudu 0,65. Samas on Vissc-

her (2008: 489) väitnud sama kordaja väärtuseks 0,8, kusjuures kordaja oli leitud Euroopa rahvastike põhjal. Erinevus võib olla tingitud sellest, et Euroopas elavate inimeste elukeskkond on sarnasem võrreldes Hiinas elavate inimeste võimalustega.

## Kokkuvõte

Inimese pikkus on tunnus, mida mõjutab palju geene ja mitmed keskkonnategurid. Antud bakalaureusetöö eesmärk oli leida pikkuse päritavuskoeffitsient erinevatel meetoditel. Kordaja hindamiseks oli kasutatud valemit, kui konkreetne sugulusside on teada, lineaarset mudelit ja kaksikute meetodit. Analüüsimisel kasutati Tartu Ülikooli Geenivaramu andmeid.

Päritavukoeffitsiendi hindamiseks on erinevaid viise ning need võivad anda ka erinevaid tulemusi. Õvede või vanem-laps paaride korral saadi pikkuse päritavuskoeffitsiendi hinnanguks ligikaudu 0,8. Samas, kui arvestatud olid kõik suguluskordajad, hinnati päritavuse suuruseks ligikaudu 0,7. Selgus, et keskkond mõjutab pikkuse varieeruvust vähesel määral. Suurem mõju on hoopis erinevatel geenidel, mida on nüüd otstarbekas leidma hakata. Kaksikute meetodil saadud tulemuste põhjal ei saanud teha üldistavaid järeldusi, sest kaksikute paare oli andmestikus vähe.

## Kirjandus

**Beebe, T. J. C, G. Rowe 2004.** An Introduction to Molecular Ecology. New York: Oxford University Press Inc.

**Gillespie, J. H. 2004.** Population Genetics A Concise Guide. Second edition. Baltimore and London: Johns Hopkins University Press.

**Haverkamp, A. 2015.** The representation of the concept of heritability in Dutch written news media. Master Thesis of Science Education and Communication Freudenthal Institute for Science and Mathematics Education, Utrecht University.

**Hedrick, P. W. 2011.** Genetics of Populations. Fourth edition. Massachusetts: Jones and Bartlett Publishers.

**Heinaru, Ain 2012.** Geneetika. Tartu: Tartu Ülikooli Kirjastus.

**Kaart, Tanel 2001.** Ülevaade geneetiliste parameetrite hindamisel kasutatavatest mudelitest. Eesti Põllumajandusülikooli Loomakasvatusteaduste instituudi teadustööd, 71: 52-66, Tartu.

[http://www.eau.ee/%7Ektanel/kaart\\_2001\\_LKI.pdf](http://www.eau.ee/%7Ektanel/kaart_2001_LKI.pdf) (vaadatud 08.12.2016)

**Kaart, Tanel 2012a.** Lineaarne geneetiline mudel, populatsiooni olulisemad geneetilised parameetrid. Kursuse Loomade aretusväärtuse hindamine ja aretusprogrammid loengumaterjal, Eesti Maaülikool.

[http://www.eau.ee/~ktanel/VL\\_0192/pt4\\_2012.pdf](http://www.eau.ee/~ktanel/VL_0192/pt4_2012.pdf) (vaadatud 25.11.2016)

**Kaart, Tanel 2012b.** Indiviidvaheline sugulus. Kursuse Loomade aretusväärtuse hindamine ja aretusprogrammid loengumaterjal, Eesti Maaülikool.

[http://www.eau.ee/~ktanel/VL\\_0192/pt3.pdf](http://www.eau.ee/~ktanel/VL_0192/pt3.pdf) (vaadatud 16.01.2017)

**Käärrik, Ene 2013.** Andmeanalüüs II, loengukonspekt. Tartu Ülikooli matemaatilise statistika instituut.

<http://dspace.ut.ee/bitstream/handle/10062/35401/AndmeanaluusII.pdf> (vaadatud 06.05.2017)

**Li jt=Li, Miao- Xin, Peng- Yuan Liu, Yu- Mei Li, Yue- Juan Qin, Yao- Zhong Liu, Hong- Wen Deng 2004.** A major gene model of adult height is suggested in Chienese. Journal of Human Genetics 49, pp. 148–153.

**NCD-RisC (NCD Risk Factor Collaboration) 2016.** A century of trends in adult human height. eLife;5:e13410

**Nyquist, W. E. 1990.** Statistical Genetics. Agronomy 615, Purdue University. Chapter 4.

<http://nitro.biosci.arizona.edu/nyquist/chap4.pdf> (vaadatud 25.11.2016)

**Oldenbroek, K, L. van der Waaij 2014** Textbook Animal Breeding and Genetics for BSc students. Centre for Genetic Resources and Animal Breeding and Genomics Group, Wageningen University and Research Centre, the Netherlands.

[https://www.wur.nl/upload\\_mm/4/d/5/3b256d1f-2ae5-4fa6-a2d7-afe355ce9870\\_Textbook%20Animal%20Breeding%20and%20Genetics-v17-20160105\\_2020.pdf](https://www.wur.nl/upload_mm/4/d/5/3b256d1f-2ae5-4fa6-a2d7-afe355ce9870_Textbook%20Animal%20Breeding%20and%20Genetics-v17-20160105_2020.pdf) (vaadatud 23.01.2017)

**Rijsdijk, F. V, P. C. Sham 2002.** Analytic approaches to twin data using structural equation models. Briefings in Bioinformatics, vol. 3, no. 2, pp. 119-133.

**Rudary, M. R 2009.** On Predictive Linear Gaussian Models. Doctor of Philosophy (Computer Science and Engineering) in The University of Michigan.

[https://deepblue.lib.umich.edu/bitstream/handle/2027.42/62307/mrudary\\_1.pdf;sequence=1](https://deepblue.lib.umich.edu/bitstream/handle/2027.42/62307/mrudary_1.pdf;sequence=1) (vaadatud 08.12.2016)

**Silventoinen jt= Silventoinen, K., S. Sammalisto, M Perola, D.I. Boomsma, et al. 2003.** Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Research*, vol. 6, no. 5, pp. 399-408.

**Sham, Pak 1998.** *Statistics in Human Genetics*. Institute of Psychiatry, London, UK: Arnold.

**Teinberg, Rein 1978.** *Põllumajandusloomade geneetika*. Tallinn: Valgus.

**Visscher, Peter M. 2008.** Sizing up human height variation. *Nature Genetics* vol. 40, no. 5, pp. 489 - 490.

# LISAD

## Lisa 1. Meeste ja naiste pikkuse sõltuvust kirjeldavate mudelite võrdlus

```
> anova(m1,m2,m3,m4,test="F")
Analysis of Deviance Table

Model 1: mehed$pikkus ~ mehed$aasta
Model 2: mehed$pikkus ~ mehed$aasta + I(mehed$aasta^2)
Model 3: mehed$pikkus ~ mehed$aasta + I(mehed$aasta^2) + I(mehed$aasta^3)
Model 4: mehed$pikkus ~ mehed$aasta + I(mehed$aasta^2) + I(mehed$aasta^3) +
I(mehed$aasta^4)
  Resid. Df Resid. Dev Df Deviance      F      Pr(>F)
1      17730      743964
2      17729      740149  1    3815.4  91.564 < 2.2e-16 ***
3      17728      738685  1    1463.8  35.130 3.141e-09 ***
4      17727      738672  1      13.5   0.323  0.5698
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

Joonis 5. Meeste pikkuse sõltuvus sünniaastast kirjeldavate mudelite võrdlus

```
> anova(n1,n2,n3,n4,test="F")
Analysis of Deviance Table

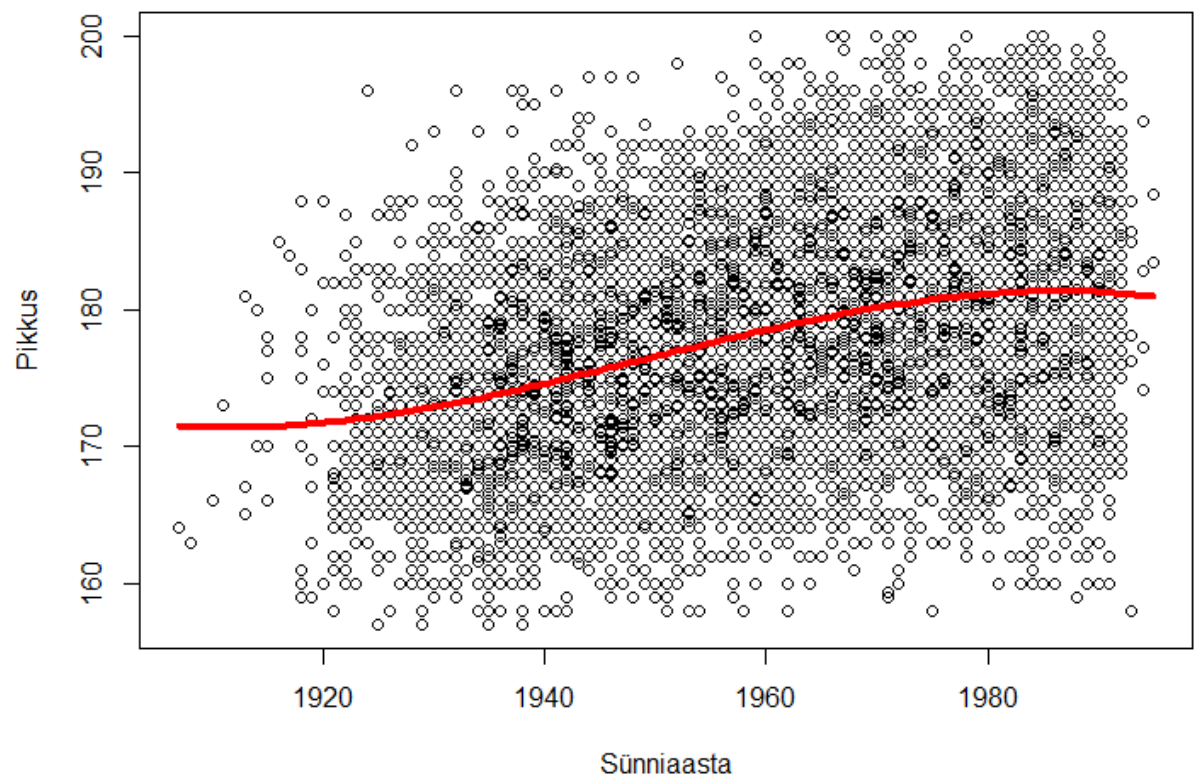
Model 1: naised$pikkus ~ naised$aasta
Model 2: naised$pikkus ~ naised$aasta + I(naised$aasta^2)
Model 3: naised$pikkus ~ naised$aasta + I(naised$aasta^2) + I(naised$aasta^3)
Model 4: naised$pikkus ~ naised$aasta + I(naised$aasta^2) + I(naised$aasta^3) +
I(naised$aasta^4)
  Resid. Df Resid. Dev Df Deviance      F      Pr(>F)
1      33865      1153938
2      33864      1151585  1    2353.28 69.2299 < 2.2e-16 ***
3      33863      1151140  1     445.08 13.0935 0.0002968 ***
4      33862      1151045  1      94.84   2.7901 0.0948587 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

Joonis 6. Naiste pikkuse sõltuvus sünniaastast kirjeldavate mudelite võrdlus

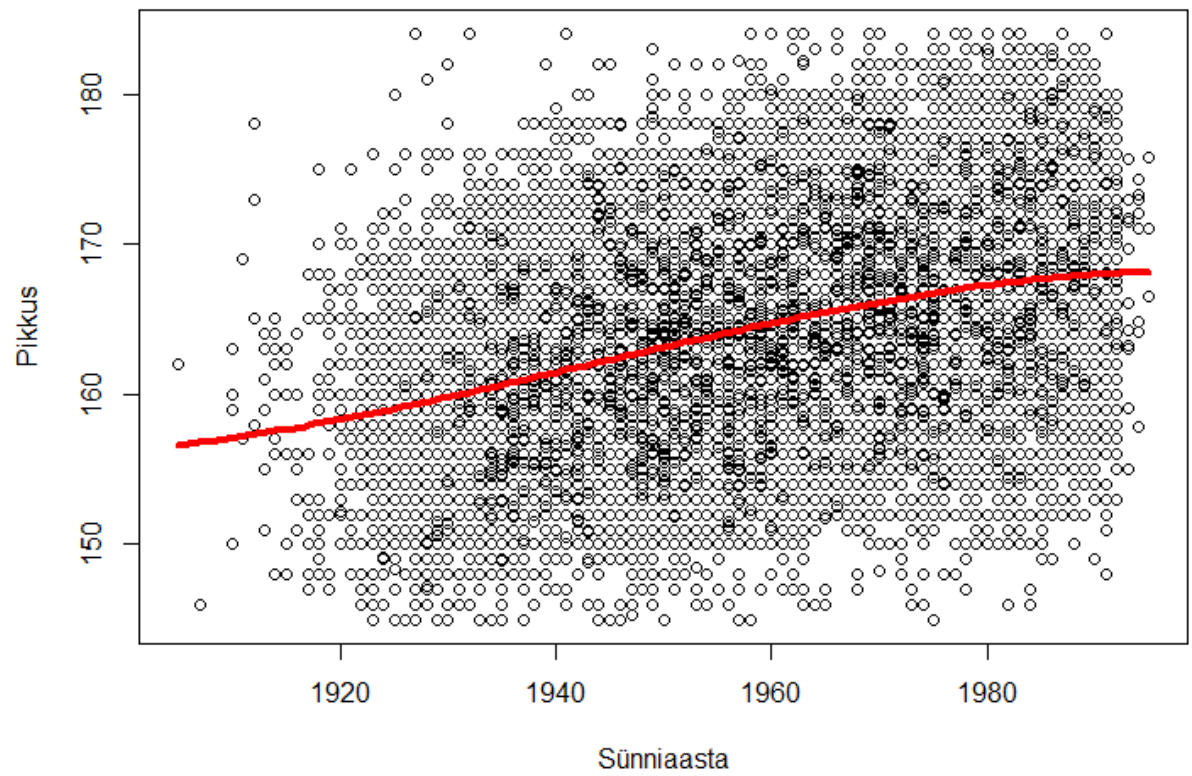
Selleks, et välja selgitada, kas kõrgemat järku mudeli kasutamine on statistiliselt oluline võrreldes madalamat järku mudeliga, tuleks kasutada F- statistikut (Käärik 2013: 98). Joonistel 5 ja 6 on välja toodud vastavalt meeste ja naiste pikkuse sõltuvust kirjeldavate mudelite võrdluste R-i väljavõtted. Neljandat järku polünoom ei ole statistiliselt oluline (p- väärtus on suurem kui 0,05), mistõttu osutub valituks kuuppolünoom.



## Lisa 2. Täiskasvanud meeste pikkuse sõltuvus sünniaastast



Lisa 3. Täiskasvanud naiste pikkuse sõltuvus sünniaastast



# Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Kaisa Komp,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Pikkuse päritavuskoefitsiendi hindamine“, mille juhendajad on Märt Möls ja Reedik Mägi,
  - 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace'is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 09.05.2017